A STANDARD OF INTERPRETATION OF NUMERICAL GRADES

LeROY D. WELD
Coe College, Cedar Rapids, Iowa

A great deal has been written regarding methods of standardizing teachers in the grading of pupils, and various schemes are in use whereby teachers are constrained to exercise arbitrary rules of judgment in order to make the results of their grading fulfil certain theoretical conditions.

We are told, for example, that, in some large American colleges and universities using forms of the so-called "Missouri system," it is expected of teachers that a specified percentage of their assigned grades shall be above 90, another specified percentage between 90 and 80, etc. (or the equivalent of these figures in terms of letters), and that the distribution thus sought is approximately that of the familiar, symmetrical "probability" or "error" law.

It is not the purpose of this paper to discuss the merits of either numerical or literal grade assignments, or of the categories "excellent," "good," "fair," etc., as against any sort of grade scale. We shall start out with the simple fact that among the great masses of our school and college teachers and pupils, the one common language in which the scholarly attainments of pupils are expressed is a scalar one, which may as well be numerical as literal or otherwise arbitrary. If we, who live in the Middle West, read in a New York magazine that a certain man entered college with an average grade of 95 in his preparatory work, we know pretty well what that means; and so it is the country over. And it will probably be a long time before the people at large will be educated to any other, radically different mode of expression.

The problem now presented is that of establishing a method whereby grades assigned by one teacher can be intelligently compared with those assigned by another, and all brought to a common standard. The writer does not believe that this can be accom-

plished by forcing teachers to conform to a theoretical system, the scientific basis of which they do not understand, and which ignores those human elements of sympathy and encouragement which make teaching the noble profession that it is. On the contrary, we propose that the teacher be let alone, left to exercise her own free mode of rating; we shall show how the grades of each teacher can be easily translated into terms of a common standard (and even into terms of the probability scale, if desired) without that teacher's knowing anything about it. Indeed it were better that she should not know her own peculiarities in this respect; for let her once be conscious that she is not quite normal in the matter of grading, and she will immediately begin, though perhaps without realizing it, to "doctor" her ratings and give constrained instead of natural judgments.

Let us first examine the ordinary percentage scale of grading and the results of its use. There can be no doubt that if it were possible to estimate accurately what is called scholarship, or proficiency, and express it in units, it would be found to have in the long run the symmetrical distribution of the theoretical "error law," like shots to right and left of a target, or the statures of people above and below the average.[1] But the fact is that teachers do not, in grading, take the same attitude toward good students and poor ones. Almost without exception, they mark the poor students higher in proportion to their attainment than they do the good students, thus revealing either the element of sympathy already referred to, or some less worthy motive, as of passing along dull pupils in order to get rid of them. This tendency has been proved in two independent ways: (1) the testimony of the teachers themselves, many of whom have been questioned on this point and have almost invariably admitted being conscious of an inclination to "shove along" the poor pupil and grade him higher than he deserves; and (2) the statistical evidence based upon a study of many thousands of grades assigned by both public-school and college teachers. This latter investigation has given some very interesting results, and is the basis of our proposed method of standardizing teachers' assigned grades.

[1] See Weld, *Theory of Errors and Least Squares*, chap. iv.

The grade lists used in this study were obtained in part from college records and in larger measure from the records of the public schools of Cedar Rapids, kindly furnished by Superintendent J. J. McConnell for the purpose, over one hundred thousand individual grades being tabulated from lists assigned by about one hundred and fifty teachers over a period of several years. The public-school grades were numerical, each grade being assigned to the nearest multiple of five; for example, the grades 73, 74, 75, 76, 77 were all called 75, while 78, 79, 80, 81, 82 were called 80, etc. The college grades were literal, each having, however, a well-understood approximate numerical significance. The work of tabulation was carried out by Mr. Leslie L. Fishwild in 1915, his summarized results being as follows:

Out of over 100,000 grades, practically none were below 50.

1 per cent were 50	13 per cent were 80
1 " " 55	13 " " 85
2 " " 60	25 " " 90
2 " " 65	23 " " 95
5 " " 70	9 " " 100
6 " " 75	

These results are shown graphically in Fig. 1, the unsymmetrical character of which is unmistakable evidence of the tendency to crowd poor students up the scale. Fig. 2 shows the normal probability distribution assumed to be ideal by the users of that system.

It is interesting to study in this manner the grades assigned by individual teachers, as their personal characteristics in grading are brought out very distinctly in this way. Some show much greater crowding than this average, some much less; occasionally a teacher will show very erratic tendencies; and very rarely one is found whose grade distribution is approximately symmetrical as theory would demand.

The writer has taken up the subject of this actual grade distribution as a mathematical problem, basing the theory upon certain very simple assumptions involving three separate personal characteristics of the individual teacher in grading, the result being a formula that agrees very closely with the statistical facts. It has been found in practice, however, that one of these characteristics,

viz., the range to which practically all the teachers' assigned grades are confined, undergoes little variation and can therefore be assumed as constant for all teachers, which leaves but two personal characteristics to be determined in order to find the type of marker to which the individual teacher belongs; as a mathematician
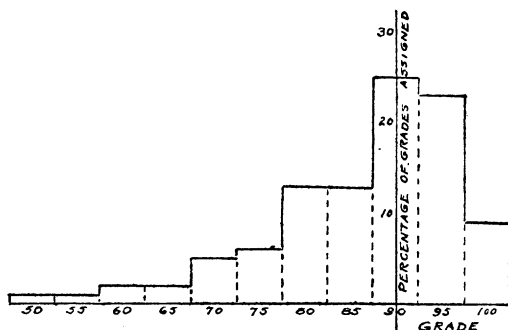
FIG. 1
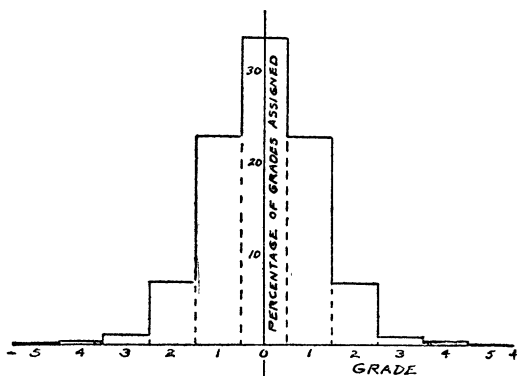
FIG. 2

would express it, only two parameters are necessary in the teachers' grade-distribution formula. This mathematical work may be published more appropriately elsewhere; but its outcome is the simple and practical method now to be presented, whereby any superintendent or principal or college registrar can determine the teacher's type of grade assignment from a single semester's grades, and be able thereafter to translate the grades given by that teacher

to a standard scale, which may be used whenever a student's actual ranking is to be determined.

For the purpose of this method the data are tabulated in a more convenient form than the foregoing. Instead of finding the percentage of a teacher's grades which are, say, 70 or 65, as in the mathematical treatment referred to, we find the percentage which are 70 *or above*, 65 *or above*, etc., and tabulate these values. It was found that of the thousands of grades examined by Mr. Fishwild, vary approximately

100 per cent were 50 or above	83 per cent were 80 or above
99 " " 55 " "	70 " " 85 " "
98 " " 60 " "	57 " " 90 " "
96 " " 65 " "	32 " " 95 " "
94 " " 70 " "	9 " " 100
89 " " 75 " "	

The passing grade being 75, 11 per cent of the grades denote failure.

While it would, of course, be desirable to make a selection from various localities, it is fair to presume that this distribution is not far from normal the country over, since the teachers assigning these grades were not by any means all of local origin. At any rate this distribution, even if not quite the average for the United States, will serve our purpose as a reference point and means of comparison. It is shown graphically by the heavy curve in Fig. 3, along with the corresponding curves for certain individual teachers, which are dotted.

Now the foundation principle of our method is that *we may expect any one large group of unselected pupils (as those handled by one teacher in a year) to have about the same actual scholarship, on the average, as any other similar large group*. This means that radical differences observed in the grade distribution of one teacher from that of another have their origin in the characteristic grading methods of the teachers themselves rather than in the pupils they handle. (If in any case there is reason to believe otherwise, due allowance should of course be made for the fact in applying the method.)

We may now proceed to classify teachers into types, according to their peculiar characteristics in grading. This may be done as

minutely and over as large a range as we think best, but the writer suggests the use of twenty types, consecutively numbered, of which the middle ones are nearest to the normal, or standard. It is believed that this number will suffice in practice. These types are identified, as before mentioned, by means of two simple charac-
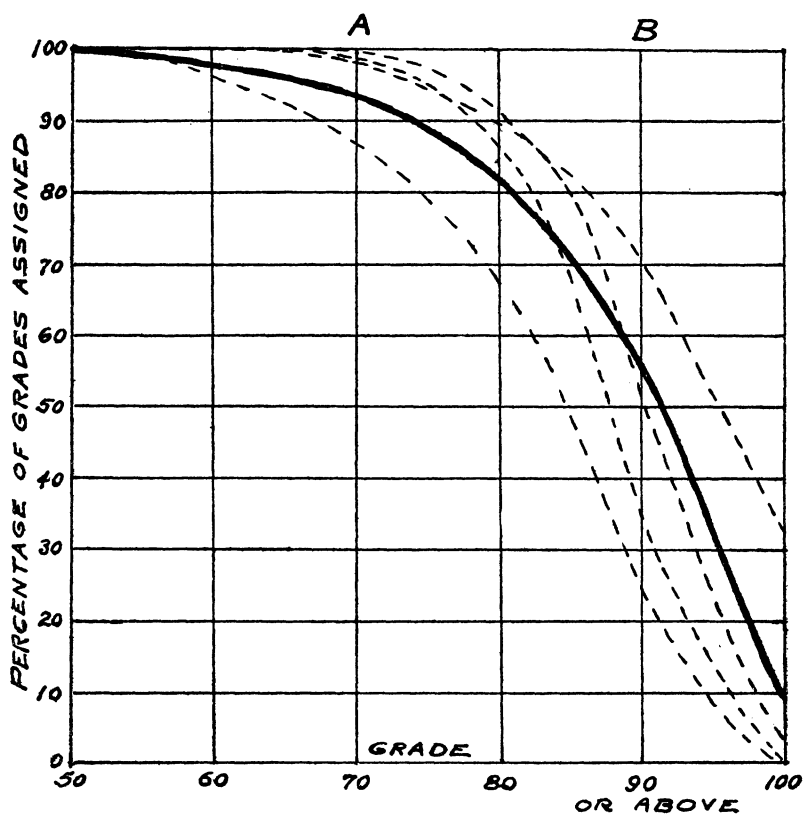
FIG. 3

teristics, for which I have selected (A) the percentage of the teacher's assigned grades that are 70 or above, and (B) the percentage that are 90 or above. (The former, A, is more important in identifying the type than the latter, B.)

For the average or standard distribution that we are using, A is 94 per cent and B is 56 per cent, which, by the way, is rather surprising when we think of it. It is an actual fact, however, that

56 per cent of all the grades examined by Mr. Fishwild were 90 or above, which exhibits more strikingly than ever the general tendency to crowd up the scale.

For individual teachers, A and B will have different sets of values, and, after considerable study of actual distributions, the twenty types shown in Table I (at end of paper) have been selected as fairly representative of the range likely to be encountered. In general the first types correspond to the consistently low markers and the last to the consistently high markers, while the middle type (11) is about normal. Provision has been made, also, for certain types that for some reason are high markers of good pupils and low markers of poor pupils, or vice versa; it is certain that such types exist. In practice a judicious combination of two types may sometimes be found satisfactory with gradual transition along the scale from one type to the other.

The next step in the development of the method was to compare the grade distributions corresponding to the respective types with the standard distribution. Familiarity with the form of the distribution curve (Fig. 3), through plotting many individual grade distributions, made it possible to trace curves with fair accuracy when only the two characteristic points at 70 and 90 were given. A number of such curves are shown dotted in Fig. 3. The grade comparison then became a simple matter. For example, it was found from the curve for Type 16 that grades of 80 or above are given by teachers of this type in about 92 per cent of their gradings, while we observe that this same percentage of the gradings of the standard marker (Type 11) are 73 or above. We may therefore conclude that the grade 80 given by a teacher of Type 16 corresponds to the grade 73 on the standard scale of marking. In a similar manner, if a teacher of Type 6 gives a pupil the grade 80, it is equivalent to 83 on the standard scale, this teacher being a low marker.

Proceeding in this manner, it has been found possible to construct an approximate table (Table II), whereby such translations may be made at a glance as soon as the type to which the teacher belongs has been decided upon. Grades intermediate between those provided for in the table can be easily interpolated.

The possession of Tables I and II, with a semester or so of grade reports from the school system, will be sufficient to enable any school superintendent or college administrator to accomplish the desired standardization and comparison. The method of procedure may be summarized, for practical use, in the form of the following directions:

1. To standardize a teacher's grading, examine a considerable number of that teacher's grades (the more the better), ascertaining (A) what percentage of them are 70 or above, and (B) what percentage are 90 or above.

2. Find from Table I the type that corresponds most nearly with these characteristics, A and B, and assign the teacher to that type (a judicious combination of types may prove satisfactory). A should have more influence than B in selecting the type.

3. The standard grade, which corresponds to any one of the teacher's assigned grades appearing in the top row of Table II, is given below it in the same column, opposite the teacher's type number.

Example: It is found that 92 per cent of the grades assigned by Miss M. are 70 or above, and 45 per cent are 90 or above. Referring to Table I, it is seen that Miss M. belongs in the neighborhood of Type 8. She is a low marker. The horizontal row opposite Type 8 in Table II now shows that Miss M.'s 50 is equivalent to a standard 54, her 55 to a standard 59, etc.

If a teacher is found whose characteristics are nowhere near being represented by any of the types, as may sometimes occur, it is evidence of some very erratic habit of grading, and more complete tabulation of the grades will be desirable. This is likely to reveal inconsistencies which can be explained only by an utter lack of system in grading or of appreciation of what grades mean; in any case that teacher will bear watching in this respect at least. Mr. Fishwild ran across two or three cases of this sort in his research. Such instances are, however, exceptional and need not interfere with the general application of the method.

The value of such information as this method furnishes need hardly be enlarged upon. An illustration is found in the problem of selecting the honor students from the members of a graduating class who have had their instruction under different groups of teachers; or again, in the investigation of complaints as to a

TABLE I

TYPES OF TEACHERS

TYPE	A Percentage 70 or Above	B Percentage 90 or Above
1.............	85	20
2.............	86	25
3.............	87	35
4.............	88	30
5.............	89	40
6.............	90	45
7.............	91	65
8.............	92	50
9.............	93	55
10.............	93	60
11.............	94	55
12.............	95	30
13.............	95	45
14.............	96	60
15.............	97	40
16.............	97	70
17.............	98	30
18.............	98	35
19.............	99	45
20.............	99	50

TABLE II

FOR THE TRANSLATION OF GRADES TO STANDARD SCALE

Type of Teacher	Grade Assigned by Teacher of Given Type										
	50	55	60	65	70	75	80	85	90	95	100
1.............	55	65	70	74	78	83	87	92	97	100	100
2.............	55	65	70	74	77	82	86	91	96	99	100
3.............	60	65	68	73	77	82	85	90	94	98	100
4.............	52	58	66	72	76	81	86	91	95	98	100
5.............	60	65	68	72	75	80	84	89	93	97	100
6.............	59	64	67	72	75	80	83	88	92	97	100
7.............	55	60	65	70	74	77	80	83	87	94	99
8.............	54	59	64	68	73	77	82	86	91	96	100
9.............	54	59	62	67	71	76	81	85	90	95	100
10.............	60	63	66	68	72	76	80	84	89	94	99
11.............	50	55	60	65	70	75	80	85	90	95	100
12.............	50	50	58	63	68	75	82	88	95	100	100
13.............	50	55	58	63	68	74	79	85	93	98	100
14.............	50	54	58	62	67	73	79	84	89	94	100
15.............	50	55	56	59	64	72	80	87	93	98	100
16.............	50	55	57	59	64	69	73	79	85	92	99
17.............	50	53	55	58	61	70	78	87	95	100	100
18.............	50	53	55	58	61	71	79	86	94	99	100
19.............	50	53	56	57	58	64	73	82	92	98	100
20.............	50	51	53	54	55	63	76	84	91	97	100

teacher's grading or of the suspicion that a teacher is being too easy, etc.

It is further to be noted that the standards of grading in different schools may be compared in exactly the same manner as those of different teachers, the process being capable of considerably greater refinement because of the larger amount of data available. This might be made use of in the rating of high schools by college-entrance boards, so that, for example, a student coming from a certain high school with an average of 87 could be considered to have an average of 85 from a standard high school.

It is the purpose of the writer, as time permits, to gather data from a wider field and by their use to improve Tables I and II, so that they may attain the greatest possible accuracy and applicability. Statistics of this kind, and suggestions of educators relative to this subject, will be appreciated. Meanwhile it is hoped that the method as presented, with the accompanying tables, will be found useful.